3D Ground-Truth Systems for Object/Human Recognition and Tracking

Afzal Godil, Roger Bostelman, Kamel Saidi, Will Shackleford, Geraldine Cheok,
Michael Shneier, and Tsai Hong

National Institute of Standards and Technology, Gaithersburg, MD, USA

Abstract

We have been researching three dimensional (3D) ground-truth systems for performance evaluation of vision and perception systems in the fields of smart manufacturing and robot safety. In this paper we first present an overview of different systems that have been used to provide ground-truth (GT) measurements and then we discuss the advantages of physically-sensed ground-truth systems for our applications. Then we discuss in detail the three ground- truth systems that we have used in our experiments: ultra wide-band, indoor GPS, and a camera-based motion capture system. Finally, we discuss three different perception-evaluation experiments where we have used these GT systems

1. Introduction

We have been researching three dimensional (3D) ground-truth systems for performance evaluation of robot perception systems in the fields of smart manufacturing and robot safety. Object recognition and localization are among the most common and challenging tasks that a robotic perception system must accomplish. These tasks are necessary to support more complex perception tasks such as identifying meaningful events and activities. In our studies, an object can be a robot, an automated guided vehicle (AGV), a person or limb, a queue of people, or any other object commonly found in an industrial environment. The goal of object recognition is to use sensed data to correctly identify objects that are present in a 3D scene. Achieving this goal is complicated because the scene could be cluttered, objects could occlude one another, and there could be illumination or viewpoint variations.

Object recognition and localization are important in many practical applications such as manufacturing automation, navigation, part inspection, and computer aided design/computer aided manufacturing (CAD/CAM), among others. Our main interest is evaluating algorithms used to recognize objects for manufacturing applications in a dynamic indoor factory environment. We emphasize the recovery of position and orientation (pose), motion, and classification of these objects so that, for example, we can determine if a person or object is moving across the scene. We are less concerned with the identification of individual people or objects. The following scenarios are those for which we would like to capture ground-truth data.

- Human and object detection and tracking from a moving platform
- Human detection and tracking for safety
- Articulated human motion tracking
- Tracking of robots and AGVs
- Human-robot collaboration

Our approach to algorithm evaluation is to compare algorithm results on a set of objects and tasks with known ground truth. The comparison is based on standardized performance metrics, such as identification accuracy, geometric position accuracy, or robustness to scene complexity. The tasks, ground-truth data, and different performance metrics should allow researchers to fully understand the strengths and limitations of different algorithms.

Typically, ground-truth measurements should be an order of magnitude more accurate than those obtained by the algorithm being evaluated. Since the algorithm will be used in a dynamic environment, its temporal resolution should be high enough to resolve the motions of the objects and eliminate motion blur. Its spatial resolution must be high enough to resolve the locations of the objects to the accuracy needed to perform the required task. There are a number of issues that need to be resolved for a successful evaluation, such as synchronization, latency issues, and time drift between the ground-truth system and the system under test. The importance of these issues varies depending on the object, the task, and the system being evaluated.

In this paper, we present a unique way of capturing 3D ground-truth data in a common world coordinate system. We first present an overview of different systems that

have been used to provide ground-truth (GT) measurements and then we discuss the advantages of sensor-based ground-truth systems for our applications. Then we discuss in detail the three ground-truth systems that we have used in our experiments. These systems included: an ultra-wide-band system, a laser-based indoor GPS system, and a camera-based motion capture system. Finally, we discuss three different perception evaluation experiments where we have used these GT systems.

2. Overview of Ground-Truth Systems

There are four main approaches to acquiring ground truth data for object recognition and tracking (for details see [24]): annotation/label-based systems, platform-based (or fixture-based) systems, physics-based simulation, and sensor-based systems.

The most popular way to create GT for object detection and tracking is by human annotation of images (including video and depth images) using annotation/label-based systems. One of the commonly used desktop tools for video annotation is VIPER-GT [1]. With this tool, users can annotate an image by drawing a bounding box around an object, indicating its identity, and providing detailed spatial and temporal information. Annotation-based approaches are typically applied to scenarios where a scene is monitored by an image or video sensor suitable for human interpretation. They have the following advantages over other approaches.

a) Complex scenes and behaviors can be annotated by hand when effective algorithms do not exist.
b) The software is often free, allowing a low-cost entry into the project.
c) The resulting annotated data support analysis by multiple groups using multiple algorithms, so repeatability is good and cross-comparisons are easily made.

Disadvantages include the labor cost of performing annotation, the variable and often unknown accuracy and reliability of the labels, and the fact that the annotations are mainly based on the images and recorded in sensor-based coordinates instead of 3D world coordinates. Also, when multi-sensor data are captured, each sensor's data are annotated independently; this can make this approach cost prohibitive and time consuming.

Platform-based systems give ground-truth for object pose by placing the object on a platform that fixes the pose in advance of a test. They usually only work for static object recognition, although a highly repeatable fixture, such as a robot, can provide dynamic poses. Physics-based simulation systems use synthetic simulated imagery based on the laws of physics. Since the environment is simulated, the exact ground-truth is

known. On the other hand, sensor data generated from a simulation do not exactly match real world sensor data because of different noise characteristics and incomplete simulation of the real world.

These factors are significant for our interest in manufacturing and robotics applications and our need to use multiple sensors and to put the resulting GT data into a single global coordinate frame. The ground-truth system that most closely meets the needs for the applications described below is a sensor-based system in which physical attributes of the object are sensed remotely and analyzed to determine its identity, location, and/or pose. Such systems use any of a number of technologies for sensing, including radio frequency, optics (photonics), acoustics (sound), and inertial sensing. Metrics for evaluating such systems include static and dynamic precision, scalability, update rate, degrees of freedom, maximum number of tracked objects, latency, work volume, range, cost, and time to identify an object [2-6]. In this paper, we review only the three sensor-based systems that we have used to obtain ground-truth measurements.

3. 3D Ground-Truth Systems

We have used the following systems to obtain ground-truth measurements in our experiments: ultra wide-band (UWB), indoor GPS, and a camera-based motion capture system. These systems are described in the following subsections and the reasons for the selection are explained according to the needs of the applications.

3.1. Ultra-Wide-band (UWB)

The UWB tracking system [12, 13] uses a collection of UWB radio receivers located around the perimeter of the test area to track multiple static and dynamic objects with credit-card size transmitter tags (see Figure 1). Each tag sends UWB pulses, which are detected by the antenna using a combination of TDOA (Time-Difference of Arrival) and AOA (Angle of Arrival) techniques to estimate the 2D or 3D position of the tags, dependent upon the test.

The UWB tracking system works in open outdoor areas or indoor areas and can see through some types of walls, though overall accuracy can vary.

(a)   (b)   (c)

Figure 1. (a) Shows a UWB receiver deployed in the field (b) shows the asset tracking system components, ultra-wideband radio frequency receiver (shown with integrated

high-gain antenna), 1 W transmitter tag, and 30 mW transmitter tag. c) Several tags are shown attached to helmets to track people in a scene.

The UWB tracking system is suited for human and vehicle tracking because of the following characteristics. UWB is based on sending ultra-short pulses [2, 3, 4, 5] over multiple bands of frequencies simultaneously. This allows them to coexist with other radio frequency (RF) signals despite their large bandwidth. They are able to differentiate the original pulses from the reflected/refracted ones because the brief time span of each pulse reduces the likelihood of overlap. Each tag and receiver has a unique identification signal and can be used in both outdoor and indoor applications. UWB is robust and provides higher precision indoor positioning compared to other wireless technologies. These systems can cover a very large area compared to other technologies and have been used successfully for human- and object-tracking applications [12, 13].

UWB systems are available commercially and have reported accuracies of 15 cm to 30 cm compared with other RF technologies whose accuracy ranges from 1 m to 3 m. They are easier to set up because they require the installation of fewer fixed sensors compared to other types of systems. UWB systems also have the advantage over optical systems that they do not require direct line of sight. The tags can, therefore, be embedded in the object being tracked, which makes it invisible to optical systems under test and thus will not affect the performance of the system under test.

We performed characterization tests on a representative UWB system in ideal conditions to determine the least possible 2D error of the system, which was measured to have an average error of 15 cm. We have used UWB to track vehicles and personnel throughout an area over 80 000 m^2 with an average error of 23 cm. The system has an update rate of 25 Hz to 50 Hz, which is sufficient to track vehicles at highway speeds [12, 13]. We have also used it to track robots through random mazes with plywood walls (non-line-of-sight) achieving similar performance. We have not been successful in tracking tags through concrete walls, but have used additional receivers in hallways to compensate during indoor building deployments. The maximum number of dynamic and static transmitter tags used simultaneously thus far has been between 15 and 30 for marking obstacles and known fiducial points to check performance. Setup time for a new test site takes about five days, including positioning, calibrating, and testing the equipment. Returning to a previously used site takes approximately two days for calibration prior to testing.

3.2. Indoor Global Positioning System (IGPS)

The iGPS system [11] shown in Figure 2 is a 3D high precision, commercial measurement system that uses stationary laser transmitters together with receivers mounted on moving or static objects to determine the poses of those objects. It is modular, suitable for large or small volumes, and can measure multiple objects simultaneously with high accuracy for both static and dynamic objects. It is used by industrial manufacturers both for positioning and tracking applications and for robot control.

Figure 2. An iGPS transmitter and two sensor receiver bars (with cables and position computation engine).

The manufacturer-specified accuracy of 3D positions measured using the iGPS is 0.25 mm and the measurement frequency is 40 Hz. A typical measurement volume based on four to eight transmitters is 1200 m^2. Detailed system analyses are presented by Schmitt et al. [7] and Mosqueira et al. [8]. Wang et al. [9] showed that the tracking accuracy is similar to the static accuracy for speeds below 10 cm/s. However, they found that as the speed of an object increases, the tracking accuracy decreases—at a speed of 1 m/s; the mean tracking error can be as high as 4 mm. In another study, Depenthal [10] showed that when tracking objects at velocities of 3 m/s, the 3D position deviation is less than 0.3 mm. Depenthal also described the experimental comparison of the dynamic tracking performance between an iGPS and a laser tracker and showed that the iGPS performed well under dynamic conditions. The iGPS, unlike the UWB system, requires line of sight to at least two transmitters to be able to make a measurement. In our human-tracking experiments, the ground-truth sensor, a pair of iGPS vector bars, is attached to the top of a hardhat worn by the human as shown in Figure 3. We have also used the iGPS to provide ground-truth measurements of AGVs (see Section 4.2).
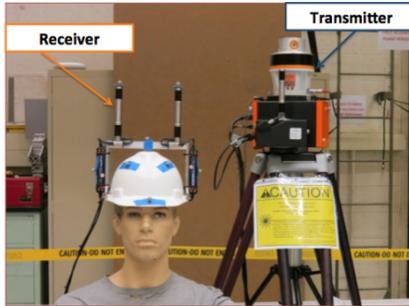
Figure 3. iGPS vector bar attached to hardhat worn by a human (represented here by a mannequin).

The iGPS sensor has a fast enough update rate to track people moving at walking or running speeds and is accurate enough to provide the necessary order of magnitude better pose measurements than most sensors used for human tracking. Its wide field of view allows a range of typical activities to be carried out by the people being tracked, and the need for only two sensors to have line of sight to the sensors at any time ensures that the system can provide data even in scenes with significant clutter.

3.3. Camera-based Motion Capture

Motion capture [21] refers to a category of methods for (1) recording the motion of objects and people, or (2) capturing the articulated motion of a whole human body and/or robotic arm. These systems are widely used to provide ground-truth for validating the performance of computer vision systems and in applications such as entertainment, sports, medical applications, and robot control. When the capture includes gestures and facial expressions or finger motions, it is sometimes referred to as performance capture. These systems can also be used to study human-robot collaboration, human-object interaction, human activity tracking, and human-human interaction applications in manufacturing environments.

The camera-based motion tracking system [15] uses a network of cameras that emit infra-red illumination. They track multiple spherical markers that reflect the illumination. When seen by more than one camera, the locations and patterns of the reflections provide pose and identity information. Balan [16] evaluated the 3D pose of human motion obtained from synchronized, multi-camera video against 3D ground-truth poses acquired with a motion capture system [17]. The synchronized video and motion capture dataset developed in [18] is the most widely used dataset. In it, actions of a single person were captured by multi-camera video together with marker-based motion capture data. The main drawback of the dataset is that there is only one person in the environment at a time, so there is no person-to-person occlusion. Other

datasets do include multiple people. One is the Utrecht Multi-Person Motion (UMPM) Benchmark [19] that includes a collection of multi-person video recordings together with ground-truth based on motion-capture data. It is intended to be used for assessing the quality of methods for pose estimation and articulated motion analysis of multiple persons using video data from single or multiple cameras.

We tested a camera-based motion capture system using a robot arm (Figure 4) in a test area of about 40 m^2, with data being collected at approximately 100 Hz from the system and at 53 Hz from the robot. The robot had previously been measured using a laser tracker to have repeatability in position of about 0.03 mm following the ISO 9283 robot performance standard. The robot moved in a path that kept the tool-control-point at a fixed height. It swept out a horizontal box with a zigzag on one side, and paused at a number of points to enable both dynamic and static performance to be measured. Because of the difference in update rates, the data from the robot and motion capture system were synchronized by interpolation. A common coordinate system was established by referencing each system to a calibration target. Thus, the results for the accuracy (or, more correctly, repeatability) of the motion capture system include calibration errors, interpolation errors, robot errors, and the system's own errors. With this caveat, we found that the system had a mean position error of 0.0140 m, with a standard deviation of 0.0116 m, when the robot was moving. For the stationary points, the mean position error was 0.013 m, with a standard deviation of 0.003 m.

Motion capture systems have the advantage of low cost relative to iGPS or UWB systems, while retaining a reasonable measurement accuracy and update rate. They are easy to set up and can cover a large area, depending on the number of cameras used.
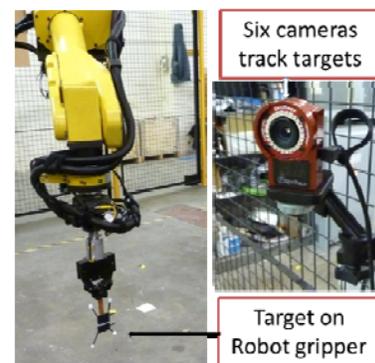

Figure 4. A robot arm with motion capture sensor and markers.

4. Perception Evaluation Using 3D Ground Truth Systems

In the following subsections we present three different perception evaluation experiments, discuss the ground-truth system used, and why the particular system was selected over other possible approaches.

4.1. Evaluating Perception Systems Mounted on a Vehicle

We conducted experiments in collaboration with the Army Research Laboratory (ARL) to evaluate six different algorithms for real-time detection and tracking of pedestrians and other objects. The algorithms used LADAR (laser detection and ranging) and video sensors mounted on a moving platform [12, 13]. For the evaluation, the moving platform was a robot vehicle equipped with two pairs of stereo cameras, two sets of imaging LADARs and two sets of 2D laser line scanners. The vehicle was driven by an operator along a straight path of approximately 240 m. Along the path were various configurations of eight moving pedestrians, four mannequins, four barrels, four cones, two trucks, two crates, seven tripods, and a number of trees. Besides variations in the complexity of the environment, the experimental variables included two vehicle speeds (30 km/h and 15 km/h) and pedestrian speeds of 1.5 m/s or 3.0 m/s.

The goal of the evaluation was to determine the performance of each of the algorithms. Because of the large area to be covered, the outdoor environment, and the relatively low position accuracy required, an UWB system was employed as the ground-truth sensor system. We developed a robust filtering algorithm to produce higher quality tracking solutions than those provided by the raw data captured by the UWB system. As described below, we developed a temporally consistent algorithm for finding the correspondence between the ground-truth data and the tracking data to improve analysis of the performance of the recognition and tracking systems. In addition, we developed a visualization tool to provide early detection of errors in data collection and to support data analysis of the test results [12].

4.1.1 Filtering the Ground-Truth Data

Filtering is a post-processing step to remove outliers and reduce error levels in the ground-truth data (for details see [12]). We first identify outliers based on the maximum conceivable speed of the tag. We then apply a polynomial least squares fit filter to a set of measurements earlier and later in time than the identified measurement. We fit a spline through the filtered points to identify the tag's position as a function of time. We interpolate the trimmed,

filtered, and splined data at timestamps obtained from each of the different algorithms. The interpolated ground-truth data are used to establish spatial and temporal correspondences.

4.1.2 Evaluation using the UWB Tracking System

The main experiment consisted of 32 runs (Figure 5). The vehicle was driven south to north on a 240-meter path. Scripted scenes with human motion, mannequins, and course clutter were sensed and interpreted and reported by the algorithms in real time. Eight humans were present in each run, four on either side of the street. Four moved parallel to the street, three at 45° to the street, and one at 90° to the street.

Figure 5. Right side of course during a run.

The performance evaluation focused on the questions of what an algorithm detected, when it detected it, and how long the detection persisted. These measures were calculated for each algorithm and in the context of the experimental factors under which the data were collected. The ground-truth system allowed the evaluation of all these questions.

Table 1 shows performance measures for the six algorithms in terms of object detections, misclassifications, and false positives over the complete set of 32 runs [12]. Entries are percentages except for the false positive entries, which report the number per run. All rows referring to objects other than people show the algorithm misclassification of these objects as humans. The evaluation performance of the six algorithms varied widely. Some systems that reported a high probability of detection also misclassified other course objects as humans. For some, the number of false positives reported was also an issue.

Table 1. Performance of six different algorithms

Object Type	Performance of Six Different Algorithms					
	Alg-1	Alg-2	Alg-3	Alg-4	Alg-5	Alg-6
Humans (%)	97.3	90.8	98.4	98.0	89.5	85.7
Mann. (%)	10.2	-	97.7	98.4	91.4	62.6
Cones (%)	0.0	-	4.7	0.0	65.6	0.0
Barrels (%)	14.1	-	54.7	70.3	89.1	0.0

Crates (%)	46.9	-	100.0	90.6	100.0	70.3
Trucks (%)	25.0	-	100.0	25.0	100.0	90.6
Tripods (%)	1.3	46.7	53.6	60.7	58.9	60.7
False Positives	29.8	77.9	155	37.3	29.8	1.3

The UWB ground-truth system was selected for this evaluation because of the large test area and the level of accuracy required. The system was used to track vehicles and personnel, including tracking vehicles moving at highway speeds. The system was also selected because part of the test area did not provide a direct line of sight between the platform and transceivers.

4.2. Evaluation of Automated Guided Vehicle (AGV) Safety Standard

NIST evaluated the performance of 2D and 3D imaging sensors on an automated guided vehicle (AGV) for safety applications. The experiments and results are presented in Bostelman et al. [20, 22]. The experiments involved comparisons of measurements of dynamic standard test pieces using sensors on an AGV, with ground-truth measurements provided by an iGPS system. The factors investigated in the experiments included the type of test piece, the type of AGV stop (controlled braking or coasting to a stop), the speeds of the test piece and AGV, the trajectory of the test piece relative to the AGV path, and operation in confined vs. open spaces. The test results will be used to develop standard test methods and to recommend improved stopping methods in an AGV safety standard [23]. Figure 6a shows the AGV instrumented with different sensors. The graph in Figure 6b shows the velocity vs. distance plot of the AGV with starting speed of 1.2 m/s after an object entered the AGV path and was detected by the onboard AGV safety sensor.
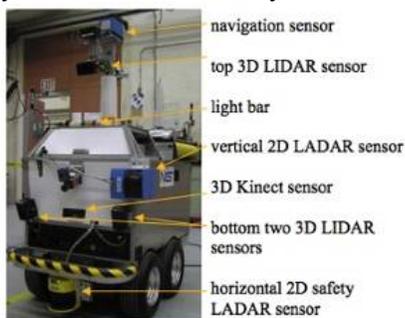

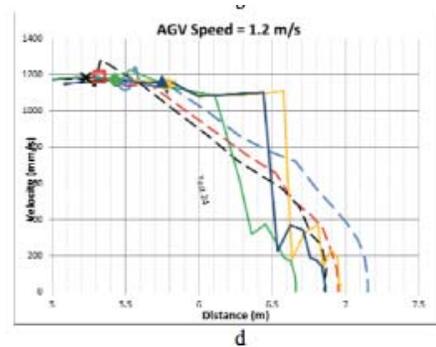Figure 6a. AGV with various onboard sensors.

Figure 6b. Velocity vs. distance plot at a starting speed of 1.2 m/s. The solid and dashed lines indicate two different types of braking – controlled braking and coasting to a stop, respectively.

AGV low-level stop braking is mandated by the safety standard, whereas controlled braking is not mandatory. Controlled braking uses the safety sensors to continually monitor areas in the AGV path beyond the low-level brake sense distance. The sensor data can be used with controlled braking to plan and execute appropriate AGV decelerations as needed when obstacles are in the vehicle path and regain speed when the path is clear. The iGPS was used to measure braking distances for both methods. The evaluation allows the AGV industry to decide which method best fits their application. The iGPS was used because this application required high accuracy measurements of relatively fast-moving objects. Even though there was significant occlusion of the iGPS receivers, there were enough transmitters (8) to ensure that accurate date could be collected.

4.3. Evaluating Perception Systems used in Workspace Situational Awareness

Next generation robotic systems will perform highly complex tasks in dynamic manufacturing environments. To be successful in performing these tasks, they need situational awareness—the ability to detect, localize, interpret, and anticipate the actions of people and objects in their environment. Prototypes of these perception algorithms are being developed, but methodologies to measure their performance do not exist. We are currently developing the metrics and methods to support the development of these methodologies, with an initial focus on the ability to detect people and objects as they move about the workspace. We will build test-beds and conduct experiments to assess the methodology. The results can be used to develop new standards that enable the use of perception systems in manufacturing applications. The first sets of experiments that are planned are on human detection and tracking for safety applications. Factors that affect the perception system's performance include

occlusion, clutter, speed of motion, and pose variation. The ground-truth will be collected by an iGPS receiver mounted on a hardhat worn by each participant as shown in Figure 3 and by a camera-based motion tracking system. The human detection and tracking experiments will be used to compare the two ground-truth systems.

We previously conducted human tracking experiments in which the system under test was a calibrated network of cameras and the ground-truth measurements were provided by the iGPS system. Figure 7 shows the setup for those human tracking experiments.
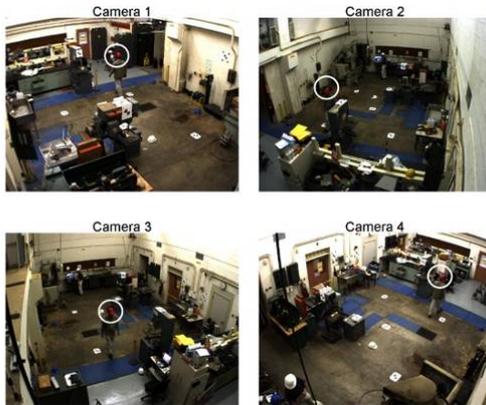

Figure 7: Different camera views during a test. Note the hardhat being tracked by the iGPS system (white circle). [14]

The localization errors between the multi-camera network and the iGPS ground-truth data are presented in Figure 8.
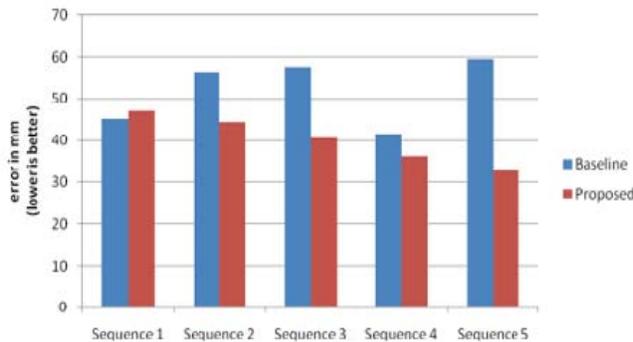

Figure 8. Graph of localization error between a base line system and a new system

5. Conclusions

This paper presented an overview of different systems that can be used to obtain ground-truth measurements. We focused on sensor-based ground-truth systems as opposed to annotation/label-based systems, platform-based systems, and physics-based simulation systems.

We discussed three systems that have been used in our experiments to obtain ground-truth measurements: UWB (ultra-wideband), indoor GPS (iGPS), and camera-based motion capture. We then provided the results of three different experiments that we have conducted and discussed why the particular ground-truth system used was better than other possible approaches.

Future work will involve establishing test-beds with systems that can provide ground-truth measurements and conducting experiments to assess the performance of perception systems. The results will provide scientific foundations for development and guidance of new standards that enable the use of perception systems in manufacturing applications.

1. Disclaimer

Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

References

[1] D. Mihalcik, and D. Doermann. The design and implementation of ViPER. Technical report, 2003.

[2] J. Torres-Solis, H. T. Falk and T. Chau. A review of indoor localization technologies: towards navigational assistance for topographical disorientation. Ambient Intelligence: pp. 51-84, Bloorview Research Institute & University of Toronto (intech Chapter), 2010.

[3] Y. Gu, A. Lo and I. Niemegeers, A survey of indoor positioning systems for wireless personal networks. Communications Surveys & Tutorials, IEEE 11, no. 1: 13-32, 2009.

[4] H. Liu, H. Darabi, P. Banerjee and J. Liu. Survey of wireless indoor positioning techniques and systems. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 37, no. 6, pp. 1067-1080, 2007.

[5] K. Al Nuaimi, and H. Kamel. A survey of indoor positioning systems and algorithms. In Innovations in Information Technology (IIT), IEEE 2011 International Conference on, pp. 185-190, 2011.

[6] R. Mautz, S. Tilch. Survey of optical indoor positioning systems. In Indoor Positioning and Indoor Navigation (IPIN), IEEE 2011 International Conference on, pp. 1-7, 2011.

[7] R. Schmitt, S., Nisch, A. Scho nberg, F. Demeester, and S. Renders. Performance evaluation of iGPS for industrial applications. In IEEE Indoor Positioning and Indoor

Navigation (IPIN), 2010 International Conference on, pp. 1-8., 2010.

[8] G. Mosqueira, J. Apetz, K. M. Santos, K., E. Villani, R. Suterio, R. and L.G. Trabasso. Analysis of the indoor GPS system as feedback for the robotic alignment of fuselages using laser radar measurements as comparison. Robotics and Computer-Integrated Manufacturing 28, no. 6: 700-709, 2012.

[9] Z. Wang, L. Mastrogiacomo, F. Franceschini, and P. Maropoulos, Experimental comparison of dynamic tracking performance of iGPS and laser tracker. The International Journal of Advanced Manufacturing Technology 56, no. 1,pp. 205-213, 2011.

[10] C. Depenthal. Path tracking with IGPS. In Indoor Positioning and Indoor Navigation (IPIN), IEEE 2010 International Conference on, pp. 1-6. ETH Zurich, Switzerland, 2010.

[11] Nikon iGPS, December 15, 2012, http://www.nikonmetrology.com/en_US/Products/Large-Volume-Applications/iGPS/iGPS

[12] B. Bodt, R. Camden, H. Scott, A. Jacoff, T. Hong, T. Chang, R. Norcross, T. Downs, and A. Virts. Performance measurements for evaluating static and dynamic multiple human detection and tracking systems in unstructured environments, PerMIS '09: Proceedings of the 9th Workshop on Performance Metrics for Intelligent Systems, 2009.

[13] B. Bodt and T. Hong. UGV Safe Operations Capstone Experiment, Army Science Conference Paper, 2010.

[14] I. Katz, K. Saidi, and A. Lytle. "Model-based 3d tracking in multiple camera views." In 27th International Symposium on Automation and Robotics in Construction, ISARC, pp. 273-279. 2010.

[15] OptiTrack motion capture system http://www.naturalpoint.com/optitrack/, December 15 2012.

[16] A. Balan, L. Sigal and M. J. Black. A Quantitative Evaluation of Video-based 3D Person Tracking, Proceedings 2nd Joint IEEE International Workshop on VS-PETS, Beijing, 2005.

[17] L. Sigal, A. Balan, and M. J. Black. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. International journal of computer vision 87, no. 1: 4-27, 2010.

[18] L. Sigal and M. J. Black. Humaneva: Synchronized video and motion capture dataset for evaluation of articulated human motion. Brown University TR 120, 2006.

[19] N. A. Van der, X. Luo, G.-J. Giezeman, R. Tan and R. C. Veltkamp. UMPM benchmark: A multi-person dataset with synchronized video and motion capture data for evaluation of articulated human motion and interaction. In Computer Vision Workshops (ICCV Workshops), 2011.

[20] R. Bostelman, W. Shackleford, G. Cheok, and R. Norcross "Standard test procedures and metrics development for automated guided vehicle safety standards." *Proceedings of the Workshop on Performance Metrics for Intelligent Systems*. ACM, 2012.

[21] T. B. Moeslund, A. Hilton, and V. Krüger. A survey of advances in vision-based human motion capture and analysis. Computer vision and image understanding 104.2: pp. 90-126, 2006.

[22] R. Bostelman, W. Shackleford, G. Cheok, K. Saidi, Safe Control of Manufacturing Vehicle Research Towards Standard Test Methods, *Proc. International Material Handling Research Colloquium*, June 25-28, 2012

[23] ANSI/ITSDF B56.5-2012, Safety Standard for Driverless, Automatic Guided Industrial Vehicles and Automated Functions of Manned Industrial Vehicles. 2012

[24] A. Godil, R. Eastman, T. Hong, Ground Truth Systems for Object Recognition and Tracking, NISTIR 7923, April 2013.

[25] S. Szabo, W. Shackleford, and R. Norcross, A Testbed for Evaluation of Speed and Separation Monitoring in a Human Robot Collaborative Environment, NISTIR 7851, March 2012.